Of Evolution, Systems and Complexity

Guillaume Beslon, Vincent F Liard, David P. Parsons, Jonathan Rouzaud-Cornabas

Of Evolution, Systems and Complexity

Guillaume Beslon, Vincent Liard, David P. Parsons, and Jonathan Rouzaud-Cornabas

Université de Lyon, INSA-Lyon, LIRIS CNRS UMR 5205, INRIA Beagle Team
56 Bd. Niels Bohr, 69 100 Villeurbanne, France.
guillaume.beslon@inria.fr

Abstract

The question of complexity in biological systems is recurrent in evolutionary biology and is central in complex systems science for obvious reasons. But this question is surprisingly overlooked by Evolutionary Systems Biology. This comes unexpected given the roots of systems biology in complex systems science but also given that a proper understanding of the origin and evolution of complexity would provide clues for a better understanding of extant biological systems. In this chapter we will explore the links between evolutionary systems biology and biological systems complexity, in terms of concepts, tools, and results. In particular, we will show how complex models can be used to explore this question and show that complexity can spontaneously accumulate even in simple conditions owing to a "complexity ratchet" fuelled by sign-epistasis.

1 Introduction

The link between evolution and complexity is as old as the evolutionary theory itself[1]. However, it is still largely controversial. There is kind of a general agreement that complexity has globally increased along evolutionary history (McShea, 1996) — although it may have decreased in some lineages, typically endosymbionts and marine cyanobacteria (Batut et al, 2014) — and that all extant organisms can be considered "complex" but the question becomes highly controversial when it comes to the origin of complexity, i.e. to its ultimate causes (Mayr, 1961).

The question of the evolutionary origin and dynamics of biological complexity should logically be a central matter of interest in Evolutionary Systems Biology (ESB). Yet, although not completely absent (Soyer and Bonhoeffer, 2006), it is surprisingly overlooked compared to questions about modularity or robustness for instance. This is surprising as a proper understanding of the evolutionary origin of biological complexity could provide valuable clues to analyse extant biological systems and help decipher the structure-function relationship in biology.

In this chapter we will first discuss the specific aspects of ESB questions with a focus on the question of the evolution of complexity (section 2). We will then discuss why and how "complex models" are required to tackle this question in particular and ESB questions in general (section 3). Finally, in section 4, we will present a series of experiments based on the Aevol model that shed a new light on the old question of the evolution of complexity.

2 Of Evolution, Systems and Complexity

The definition of Evolutionary Systems Biology is a recurrent matter of discussions within the community (Soyer and O'Malley, 2013; Loewe, 2016). But although defining precisely ESB and its relations to evolutionary biology on one side and systems biology on the other side could be an important matter for science policy, it is of low practical importance in performing science itself. Indeed, a field can afford to be fuzzily defined. What is important is to identify a coherent corpus of concepts, questions, tools and methods that are shared within a given scientific community. Ultimately, the latter determines what is considered a valid result within the boundaries of a given science and what is not.

[1]" [...] if we know of a long series of gradations in complexity, each good for its possessor, then, under changing conditions of life, there is no logical impossibility in the acquirement of any conceivable degree of perfection through natural selection." (Darwin, 1859), page 204, Chap. VI.

Linked with the attempts to define ESB, one could typically discuss whether ESB embraces evolutionary biology and systems biology, whether ESB addresses those questions that are shared by both fields or whether ESB comes with its own set of questions. However, if, as suggested above, we consider this problem at the level of concepts, questions and tools, it immediately appears that, in terms of concepts, ESB unifies both fields while, in terms of questions, ESB appears more as an intersection — if not as a disjunction — of evolutionary biology and systems biology.

Logically, ESB concepts include many concepts that originate, on the one side, from evolutionary biology (typically *Selection*, *Fitness*, *Population*, *Mutations*, *Drift*, *Epistasis*, *Fitness Landscape*, *Genotype-to-Phenotype (G2P) map*...) and, on the other side, from systems biology (*Regulation Networks*, *Metabolic Networks*, *Pathways*, *Motifs*, *Architecture*...). Given the strong roots of systems biology itself in complex systems science, it is not surprising to find also concepts initially formulated in this domain, typically: *Multiscale Systems*, *Complex Networks*, *Self-Organisation*, *Modularity*, *Feedback*... Some concepts, initially formulated in one field, have strongly benefited from the interaction with the other and have fructified therein. Typically, the concepts of *Robustness* and *Evolvability* that initially emerged in the context of complex systems science (Simon, 1962) are now fully integrated in the corpus of evolutionary biology (Wilke et al, 2001; De Visser et al, 2003; Pigliucci, 2008; Woods et al, 2011). Finally, two concepts have acquired a specific status within ESB: *fitness landscapes* and *genotype-to-phenotype maps*. Initially formulated within the context of evolutionary biology and genetics (Wright, 1932; Alberch, 1991), they quickly became central in ESB, probably because they directly match systems biology and complex systems science concepts (respectively *energy landscapes* and *multiscale systems*), hence enabling direct exchanges between both fields.

When it comes to questions, the integrative trend that we observe for concepts seems to be reversed and ESB questions seems more to intersect than to unify evolutionary and systems biology. In fact, one could easily understand that ESB, that has emerged as a scientific field after both its "parent fields", cannot address the same questions they do. To be recognised as an independent scientific field, ESB must identify its own, independent, corpus of questions: questions that cannot be answered (or even addressed) within the fields of evolutionary biology or systems biology alone but that, on the opposite, require to link both fields. Given this, it appears that the questions addressed by ESB belong to two families:

(*i.*) *What kind of systems are likely to result from a given evolutionary process?*

(*ii.*) *Knowing its properties, how is a given system likely to evolve?*

Of course, both kinds of questions are very general and lead to more specific ones (*e.g.*, How may a change of the mutation rate/mutational pattern/population size/... impact a given characteristic of the system?) but all ESB questions ultimately belong to one of these two families. Importantly, both families of questions cannot be addressed by evolutionary biology or systems biology alone as they require manipulating concepts originating from both fields. Typical examples are the influence of a system's robustness on its evolution (Wilke et al, 2001) and the influence of evolutionary conditions on a system's modularity (Kashtan and Alon, 2005) or evolvability (Crombach and Hogeweg, 2008).

Surprisingly, among the properties of biological systems, *complexity* has received relatively little attention within the field of ESB. This is surprising because the question of the origin of biological complexity is a recurrent source of debate in evolutionary biology (Gould, 1996; McShea, 1996; Dawkins, 1997), because it is at the heart of several unsolved questions, among which the well known C-Value enigma (Thomas Jr, 1971; Elliott and Gregory, 2015), and because it clearly requires integrating concepts from evolutionary biology, complex systems sciences and systems biology.

It is difficult to identify the reasons for this lack of interest, but it is worth noting that this question is also absent from the systems biology corpus. This may have two explanations. First, systems biology focuses on extant organisms which are all considered complex. Considering that biological epistemology is rooted in classification and comparison, complexity can easily be ignored as a question. Second, and more importantly, from the outside of evolutionary biology, complex biological systems are generally implicitly believed to be produced by selection (Lukeš et al, 2011). Complexity therefore seems a non-question and systems biology focuses mainly on the *function* of pathways, networks or elements, following the naive idea that they are there because they have been selected for while they could very well flourish by the sake of random drift rather than by selective necessity (Lynch, 2007).

Finally, and directly related to the questions, is the matter of tools and methods. Methods change very quickly in science and evolutionary biology, systems biology and evolutionary systems biology have all followed the "big-data bioinformatics" trend (Greene et al, 2014), accumulating deep-sequencing data on an ever-increasing number of organisms, systems and conditions. Now, accumulating data is of low interest if this data is not to be integrated into a coherent explanation framework — an unfortunate trend in the current "big-data-machine-learning" era (Pigliucci, 2009). On that matter, both evolutionary biology and systems biology share the same tradition of explanatory modelling

to seek for unifying principles. Typical explanatory models comprise fitness landscapes and population genetics (in evolutionary biology) or complex networks and dynamical systems (in systems biology). But these tools only integrate concepts from their origin field. Addressing ESB questions requires tools that integrate concepts originating from both fields (e.g., tools integrating concepts such as fitness or drift *and* concepts such as modularity or complexity). In other words, we need tools to observe how systems evolve when systems biology provides us with tools to understand systems (independently from their evolution) and evolutionary biology provides us with tools to understand the evolution of independent items (e.g. genes or traits) but lacks tools to understand the evolution of systems integrating a large number of interacting elements. What ESB requires is a set of tools enabling to observe and analyse the evolution of systemic properties within the framework of Darwinian evolution.

3 Of Complex Evolution Models

Darwinian evolution has the advantage of being relatively simple to reproduce artificially. Indeed, since the emergence of Evolutionary Algorithms (EAs) in the 1950's and 1960's, it has been shown that virtually any data structure (variables, sets, vectors, matrices, programs, networks, trees...) can evolve *in silico* as long as it is subjected to a selection process and to a replication-with-variation process. While EAs use this capacity for optimisation purposes, it can also be used to design models of evolution. Now, since there is no limit to the complexity of the data structures that can evolve within such a framework, one can design data structures with specific systemic properties and experimentally — but computationally — study how these properties evolve under various evolutionary constraints (Fig. 1). This approach emerged within the field of artificial life (Ray, 1991; O'Neill, 2003) and differ from other modelling approaches in evolutionary biology by the use of "complex models", i.e. models in which the "digital organisms" (Adami, 2006) are purposefully complex, redundant and degenerated. This introduces many degrees of freedom in the organisms' genotype-to-phenotype map, hence allowing to study the evolution of organisms' structures in parallel with the evolution of their phenotypes and fitnesses. We will hereby call this approach "*in silico* Experimental Evolution" (ISEE) to emphasize its methodological similarities with *in vivo* Experimental Evolution (Hindré et al, 2012; Batut et al, 2014).

ISEE methodology is in some way closer to *in vivo* experimental evolutionary assays than to classical population genetics models. Indeed, after the initial design of the data structure (see Fig. 1) and of the experimental conditions, ISEE practitioners let populations evolve in controlled conditions. Then, exactly as experimentalists would, they analyse *a posteriori* the winning lineages (i.e. the final data structure's systemic properties) and the underlying evolutionary dynamics (including the sequence of fixed mutational events) to relate the digital organisms characteristics with the evolutionary conditions. Although ISEE is of course limited by its use of digital rather than biological organisms, it offers numerous advantages compared to *in vivo* assays. One is of course time. ISEE offers the possibility to simulate the evolution of hundreds of populations for hundred of thousands of generations. But more than that, it allows for perfect fossil records of the simulations (Adami, 2006; Hindré et al, 2012), hence making it possible to decipher contingent events from deterministic ones, a classical issue in evolution. Last but not least, it allows for "impossible experiments" (O'Neill, 2003), i.e. experiments that would be infeasible *in vivo*, either because the experimental conditions would be impossible to set-up or because too many confounding factors would interact within biological organisms.

Compared to other modelling approaches in Systems Biology, the interest of ISEE is straightforward: it allows for the simulation of the evolution of any systems biology model provided that one can define mutation and selection operators on that model (the former is generally relatively straightforward but the latter may be tricky as it requires to be able to estimate the fitness corresponding to any parameter set). Compared to other modelling approaches in Evolutionary Biology (typically population genetics and quantitative genetics), ISEE is particularly suited to study the evolution of systems. Indeed, there is virtually no limit to the complexity of the data structures that could evolve within a computer, as long as variation and selection operators can be defined for these data-structures. In the context of ESB, this allows to use Data-structures in which systemic properties (S in Fig. 1) can vary by mutations and to observe how these properties evolve under different conditions. As long as the underlying data-structure is redundant (formally if the application DS $\rightarrow f$ is non-injective in Fig. 1), the systemic properties S can evolve independently from the fitness of organisms f. It is them possible to study the indirect interplay between evolution and these systemic properties. Indeed, in the last twenty years, many different data structures, including programs (Wilke et al, 2001; Adami, 2006), networks (Kashtan and Alon, 2005; Espinosa-Soto and Wagner, 2010), differential equation systems (Soyer and Bonhoeffer, 2006), etc., have been used to study properties such as robustness (Wilke et al, 2001), evolvability (Crombach and Hogeweg, 2008) or modularity (Kashtan and Alon, 2005; Espinosa-Soto and Wagner,

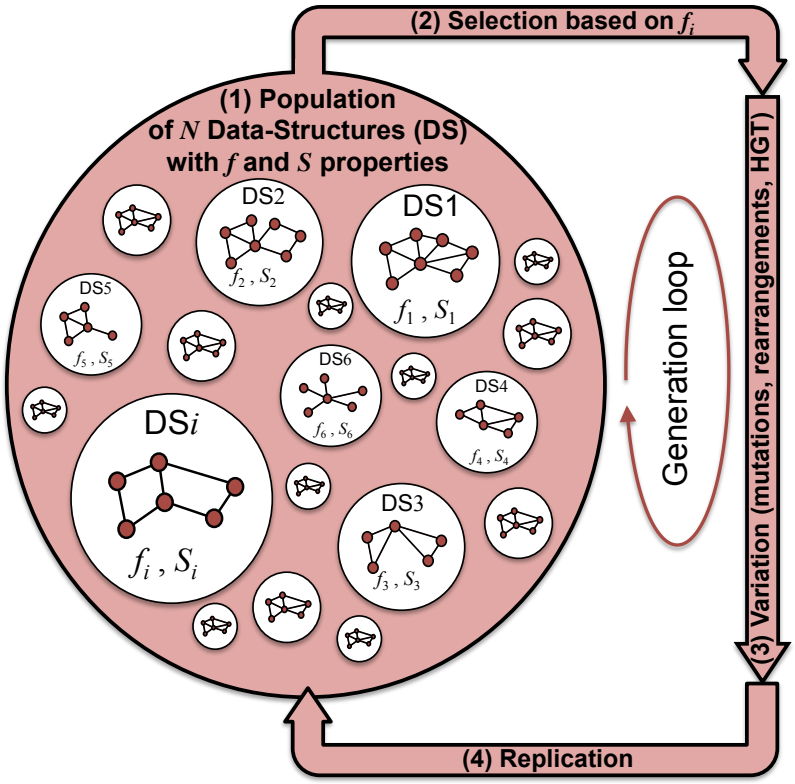

Figure 1: Sketch of *in silico* experimental evolution. The model uses a population (1) of N Data Structures (DS1, DS2,... DSi,... DSN), each associated to a computable fitness f_i and quantifiable systemic properties S_i. DSs are submitted to a generational loop composed of a selection process (2) based on f_i (but generally independent of S_i), a mutation process (3) that can modify the DSs (hence modifying f_i and S_i) and a replication process (4) that populates the forthcoming generation. The population is let to evolve under monitoring, usually for thousands of generations, before the trajectories $f_i(t)$ and $S_i(t)$ are analysed.

2010; Clune et al, 2013). However, surprisingly, only few ISEE assays have tackled the question of complexity (Adami et al, 2000; Lenski et al, 2003; Soyer and Bonhoeffer, 2006). Apart from the aforementioned reasons, another issue limits the capacity to study complexity with digital models. Indeed, most ISEE data structures have a fixed complexity. For instance, in most network models, the strength of the connections can evolve whilst the size of the network cannot (number of nodes, number of connections). In other words, one could state that most complex models are not complex enough to tackle the question of complexity.

4 Of Evolution of Complexity

4.1 Introduction

The evolutionary origin of complexity is a historical source of controversy. Basically, there are two main theoretical bodies, each emphasizing one of the two main engines of evolution: variation and selection. A naive (but common) interpretation of Darwininan theory states that selection is the driving force at the origin of extant structures. In this view, extant complexity is "naturally" due to selection, which can act through several mechanisms (Lukeš et al, 2011): Complexity could be selected simply because complex organisms intrinsically have a higher fitness or because complex environments select for complex organisms, directly (Albantakis et al, 2014) or through interactions with other species or mates (Zaman et al, 2014); because complex organisms are more robust or more evolvable than simple ones (Soyer and Bonhoeffer, 2006); because multi-part systems require complex regulation mechanisms (Maslov et al, 2009) or because new genes are recruited to compensate for the negative pleiotropic effects previous mutations brought

about (Pavlicev and Wagner, 2012). In contrast, according to neutralist theories, complexity increases by the action of random variations that spontaneously accumulate complexity, at least in some lineages. For instance, in S. J. Gould's "drunkard's walk" model, variation disperses lineages in the space of complexity levels, hence resulting in the emergence of lineages with ever increasing complexities (Gould, 1996). In the "Zero-Force Evolutionary Law" (McShea and Brandon, 2010), variation disperses redundant components of the evolving system, hence increasing its complexity. Finally, in "Constructive Neutral Evolution" (Lukeš et al, 2011), complexity arises because mutations create dependencies in multi-components systems. So, while the precise mechanism by which variation may increase complexity depends on the authors, all neutral models agree on the idea that the main driving force is random drift.

There are many reasons why studying the evolution of complexity is difficult. First, definitions and measures of complexity are not firmly established, especially for biological systems (Adami, 2002). Second, biological systems are multi-scale systems and complexity can simultaneously rise and fall on different scales, as exemplified by the C-value enigma (Elliott and Gregory, 2015), by the strong streamlining of obligate bacterial symbionts (Moran, 2007) and by "major transitions" (Maynard Smith and Szathmary, 1997). Finally, another difficulty is the lack of experimental tools. Indeed, most of the above-mentioned hypotheses are based on thought experiments and/or on extant organisms without any possibility to observe the transition between the emergent life, supposedly "simple", and extant complex life forms.

By using complex models, ISEE can overcome these difficulties. Complexity is easier to define and measure on models than it is on real organisms and, in a simulation, complexity can be monitored all along the evolutionary process on perfect fossil records. Moreover complex models can integrate different scales and complexity can be quantified simultaneously at these different scales. Last but not least, ISEE makes it possible to perform "impossible experiments" (O'Neill, 2003) in which multi-scale organisms evolve in environments which are more or less demanding in terms of complexity. Indeed, one can define environments in which simple organisms can easily thrive (at least as easily as complex organisms) and environments in which complex organisms are likely to have a better fitness than simple ones. Then, by comparing the evolutionary outcomes in these two conditions, it is possible to decipher the relative effect of selective and neutralist forces on the dynamics of complexity.

We used the Aevol *in silico* experimental evolution platform to implement this research program. We let populations of initially simple individuals evolve in two different environments: a simple one and a complex one. The analysis of the winning lineages in the two contexts revealed that complexity evolves in both contexts and that complex organisms are not more complex in demanding environments. Moreover in both environments complexity increases are driven by selection, although in simple environments, simple organisms are far fitter than complex ones. These seemingly antagonistic results together show that complexity is driven by a ratchet mechanism, powered by selection but clicking in a direction opposite to the long range selection gradient.

4.2 The Aevol model

Aevol (`https://www.aevol.fr`) is an *in silico* experimental evolution platform developed by the Inria Beagle team. This chapter is not intended to promote Aevol, and since the platform has been described in numerous publications (Knibbe et al, 2007; Beslon et al, 2010; Batut et al, 2014; Rutten et al, 2019; Liard et al, 2020), we will confine ourselves to its core principles and focus on the structure of the information coding scheme as it is of utmost significance in our experiments.

The rationale of Aevol is that the structure of the fitness landscape of an organism is likely to be strongly determined by the structure of the biological information coding of this organism (i.e. by the structure of its genotype-to-phenotype (G2P) map) and by the variety of mutational operators that act on its genome. That is why, Aevol uses a data structure that mimics precisely the biological genomic structure and that is decoded through a bio-like G2P map (Fig. 2, parts A to D). "Organisms" are then embedded in an evolutionary loop that includes classical selection operators and a large variety of mutations operators including structural ones (Fig. 2 parts F to H).

The core principles of Aevol make it ideally suited to study the evolution of biological complexity. Aevol is a multi-scale model. As such, complexity can be monitored at different levels (genomic, proteomic and phenotypic) and the set of mutational operators includes large-scale chromosomal rearrangements (including duplications and deletions). Hence, genomic complexity can vary by gene duplication-divergence, possibly driving complexity to higher levels. Finally, similarly to "real" biological G2P mappings, Aevol's mapping is redundant meaning that complexity can evolve partly independently at the different levels.

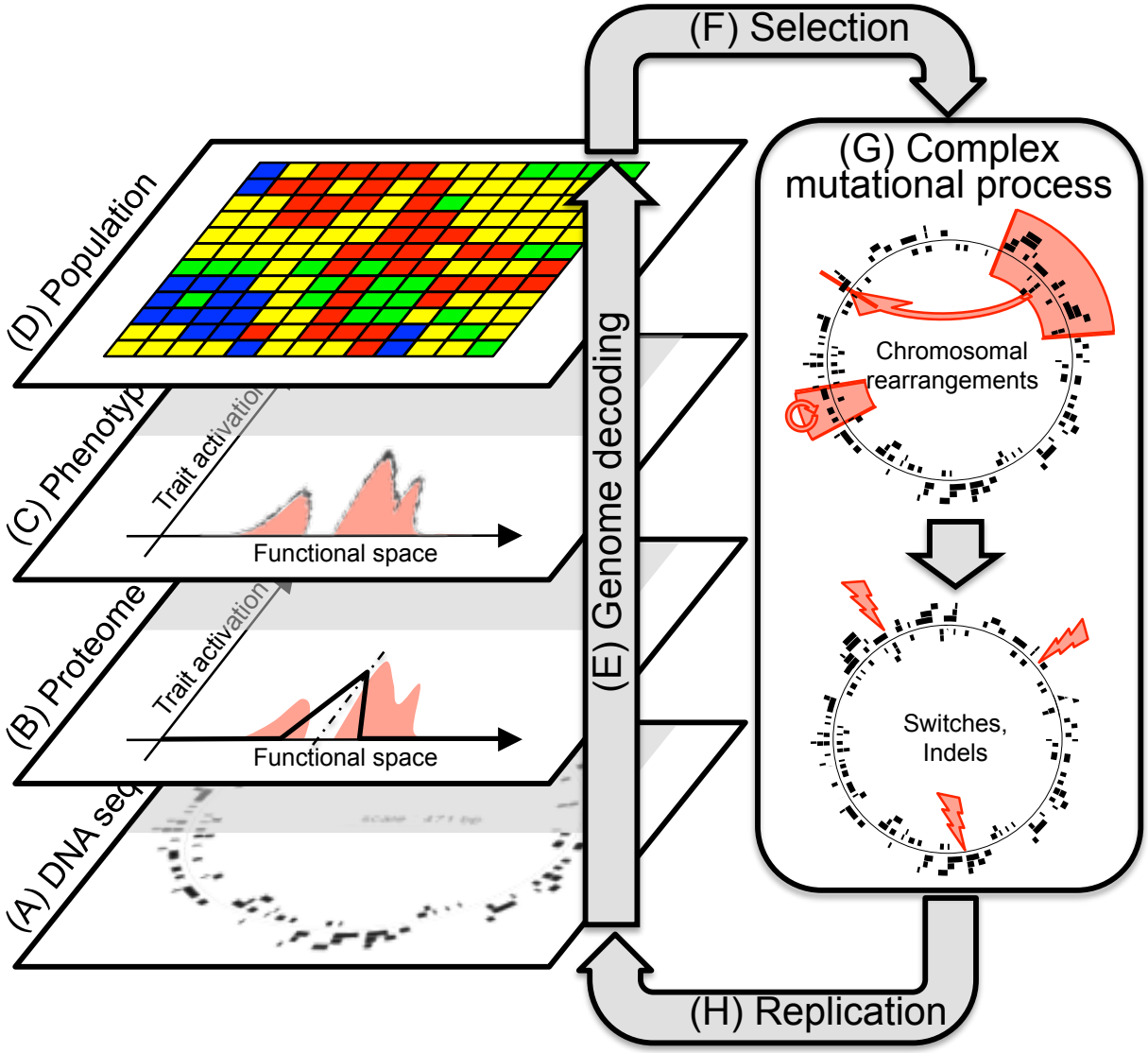

Figure 2: The Aevol model. Genomes (A) contain genes decoded into proteins. Functional levels (proteome and phenotypes) use a mathematical abstraction where proteins are represented by triangular functions (B) and phenotypes are computed as the sum of all the proteins' functions (C). Fitness is computed through a curve-fitting task, the closer the phenotype from a "phenotypic target" (in red on parts B and C) the fitter the organism. Aevol is based on a generational loop with selection (F), a complex mutational process including rearrangements and mutations (G) and replication (H).

4.2.1 Information coding in Aevol

In Aevol, each individual owns a genome containing its heritable information (Fig. 2.A). The genome is a binary double-stranded sequence that is decoded in two steps: *Transcription* and *Translation*. Transcription relies on consensus signals (promoters) and hairpin-like structures (terminators) for transcription initiation and termination respectively. Translation involves consensus ribosome binding sites and an artificial genetic code based on triplet codons (including Start and Stop) which is used to compute the protein sequences. Importantly, these processes introduce redundancy and degeneracy: complex genomes can encode for simple proteomes (e.g. if all genes have the same sequence) and complex proteomes can be encoded on compact sequences (if genes share sequences through e.g. overlapping).

Given the sequence of a protein, Aevol computes its functional contribution. Now, although mimicking biological processes at the sequence level is feasible, it is impossible to compute the function of a protein from its primary structure in a realistic way. That is why Aevol uses an abstract mathematical formalism to describe the functional levels (proteins and phenotype). In Aevol all functions are expressed in a one-dimensional continuous "functional space" (more precisely on the $[0, 1]$ interval) by an activation value in the $[-1, 1]$ interval (upper and lower bounds corresponding to maximum activation and maximum inhibition respectively). In this space, proteins are described as triangle-shaped functions (Fig. 2.B), themselves described by three parameters (mean m, height h and half-width w) computed from three interlaced variable-length binary codes in the primary structure of the protein. Once all kernels have been computed from the protein set, they are summed to compute the phenotype (Fig. 2.C).

Finally, in Aevol, the fitness is computed as the exponential of the difference between the phenotypic function and a "phenotypic target" indirectly representing the abiotic conditions the organisms evolve in. Classically in Aevol the target function is defined by a sum of Gaussians, hence requiring a virtually infinite number of protein-triangles to be perfectly fitted.

4.3 Designing an impossible experiment

The abstract mathematical formalism used to model functional levels in Aevol makes it possible to design experiments allowing to quantify the relative contribution of neutral and selective forces in the evolution of complexity. Indeed, it is difficult to quantify these relative contributions in a complex environment where they are both expected to increase complexity. Now, if one can let initially simple organisms evolve in a simple environment, then, selective forces are supposedly inactive and the evolutionary outcome shall only reflect neutral forces contribution.

Using Aevol, we can easily design such an "impossible experiment": given Aevol's G2P map, we can design two kinds of environments. As stated above, in Aevol, genes are decoded into triangular kernel functions and the sum of these kernels gives the organism's phenotype. As a result, because triangular phenotypic targets have the same shape as protein kernel functions, they don't require a complex proteome structure to be fitted (but they *can* be fitted by a complex proteome). On the opposite, Gaussian-shaped functions are impossible to fit with a finite number of protein-triangles. We used this property to design two phenotypic target functions, one simple, the other complex (Fig. 3). Then, we sampled random genomes to find "trivial" organisms with only one gene and let these initially trivial organisms evolve in both environments to quantify the final levels of complexity. Finally, since the mutation rate is also likely to influence complexity (Knibbe et al, 2007), we tested three different mutation rates: $\mu = 10^{-4}$, 10^{-5}, and 10^{-6} mut.bp^{-1}.gen^{-1}. We simulated 100 independent evolution threads per condition with a constant population size (1,024 individuals). Simulations were seeded with clonal populations of simple individuals with a single gene generated by random sampling of 5,000 bp genomes. All simulations lasted 270,000 generations.

In silico experimental evolution makes it possible to store perfect fossil records. Hence, once the evolutionary runs are finished, one can take any organism at any generation and retrieve its lineage, including fixed mutations. Given an organism one can also measure its characteristics, including its complexity, robustness and evolvability[2]. Here we measured fitness, complexity, robustness and evolvability of the ancestors of the best final organism (at generation 270,000) from generation 0 to 250,000 (the last 20,000 generations being ignored as ancestors cannot be considered to be fixed when they are too close to the final generation).

Following Adami (2002), we considered complexity as the quantity of information an evolving system integrates from its environment. Here we quantified complexity at two levels: the sequence level and the functional level. At the sequence level, *Genomic complexity* (C_G) is directly measured by the number of "essential" base pairs on the genome

[2]In Aevol, robustness and evolvability are estimated by Monte Carlo sampling. 10,000,000 offspring of a given individual are generated. Robustness is estimated by the fraction of neutral offspring and evolvability is estimated by the mathematical expectation of fitness improvement on the forthcoming generation.

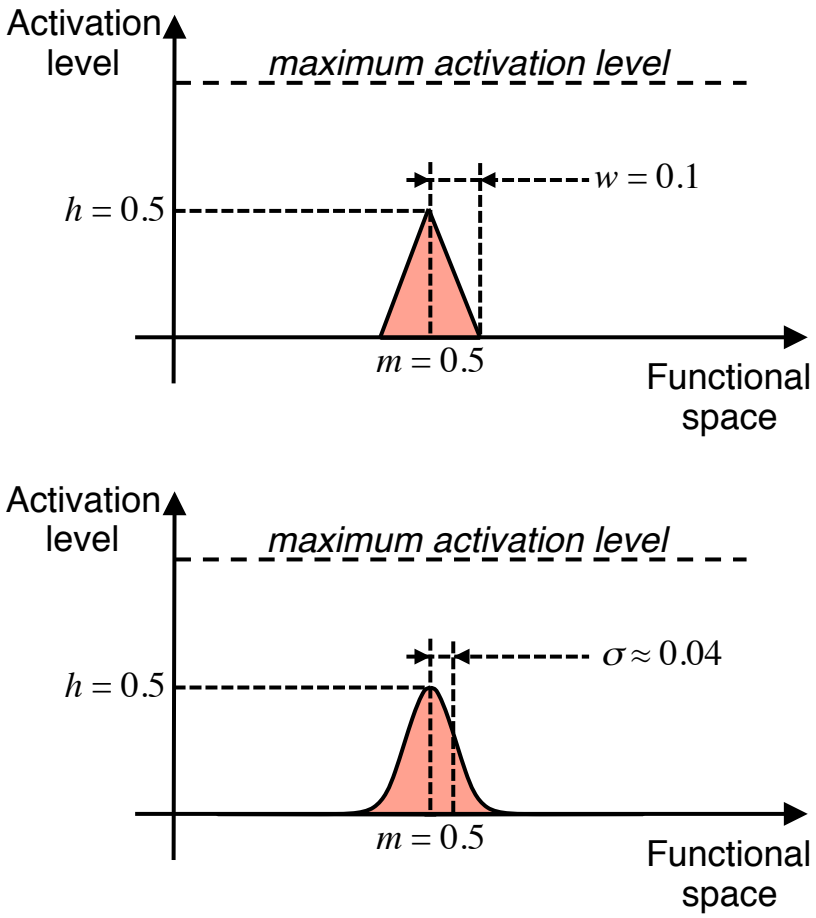

Figure 3: The two phenotypic targets used in our experiments. Top: The simple target is an isoceles triangle of mean $m = 0.5$, half-width $w = 0.1$ and height $h = 0.5$. This shape can be perfectly fitted by a single protein-triangle (see section 4.2). Bottom: The complex target is a Gaussian-shaped curve of mean $m = 0.5$, standard-deviation $\sigma \approx 0.03989$, and a maximum value $h = 0.5$. Given the proteins model in Aevol, this target cannot be perfectly fitted with a finite number of proteins.

(i.e. base pairs which, if mutated, would change the phenotype of the organism). Note that it may be very different from the genome size since the genome can accumulate non-coding sequences. It can also be shorter than the functional information stored on the genome since genes can share sequences by means of overlapping (see Fig. 4.B). Measuring *functional complexity* C_P is not as straightforward. At the phenotypic level, complexity is directly driven by selection that imposes that the phenotype fits the phenotypic target. Now, the phenotypic function can result from the sum of a variable number of protein-kernel function. However, simply counting the proteins would overestimate complexity as two proteins can have different sequences but the same function. Hence, we used a finer description of the proteomic information to measure the functional complexity C_P: C_P is defined as the number of parameter values used to encode the protein set (i.e. the number of different m, different w and different h values in the kernel set).

To study the long-term fate of simple vs. complex organisms, we also defined a qualitative classification procedure. A trivial option would have been to define a threshold on the quantitative measures but this would be arbitrary. Hence, we classified organisms according to their functional structure: In Aevol, if all the proteins of an organism have the same m and w (i.e., all the proteins have the same function, possibly with different levels of activity h), then their functions produce a triangular phenotype with the same characteristics. We used this property to define two classes. Organisms are called "simple" if all their proteins have the same function (in mathematical terms, if the sum of all kernel functions is a kernel). Importantly simple organisms may contain many different proteins differing in their efficiency h. Hence, simple organisms can have variable levels of genomic (C_G) and functional (C_P) complexities. Fig. 4 shows examples of simple and complex organisms.

4.4 Results: The complexity ratchet

Among the 300 simulations that evolved in a complex environment, 298 were classified as "complex" at generation 250,000. However, strikingly, only 71 of the 300 simulations that evolved in a simple environment are "simple" at generation 250,000, the remaining 229 being "complex". Even more strikingly, a comparison of the mean fitness of the simple and complex organisms (in a simple environment) shows that simple organisms are far fitter with a mean fitness: $\overline{f_{\text{simple}}} = 0.99 \pm 0.008$ (the maximum fitness in Aevol being: $f_{\max} = 1$) than the complex ones ($\overline{f_{\text{complex}}} = 0.42 \pm 0.32$). This shows that in a simple environment complex organisms have no direct selective advantage over simple ones. We also verified that they have no robustness and evolvability advantage (simple organisms being actually much more robust that complex ones).

A natural interpretation of these results is that complexity is a transitory state and that complex organisms will become simpler as their fitnesses improve. However, this is firmly contradicted by the dynamics of simple vs. complex individuals. First, simples' fitnesses grow very quickly at the beginning of the simulations to reach their maximum in a few thousands of generations (often less than one thousands). Second, among the 236 individuals that were complex at generation 10,000, 227 were still complex at generation 250,000. This shows that complexity (or simplicity) belongs to organisms' identities and that, once fixed at the very beginning of the simulations, hardly can it change thereafter.

When comparing the final complexity levels in the simple and complex environments, we found no significant difference. Together with the previous observation that, in a simple environment, complexity accumulates despite the fitness advantage of simple organisms, this suggests that complexity in driven by a strong neutral process, strong enough to overcome the large difference of fitness between simple and complex organisms in the simple environment (see above). However, in both environments, results show that a high mutation rate strongly limits complexity at both the genomic and functional levels, which is not consistent with a neutral process as diffusion is likely to occur at a faster pace with a high mutation rate. Moreover, when relaxing the selection pressure (i.e. letting the populations evolve further without selection), we observed that complexity quickly drops to zero, hence invalidating neutralist hypotheses.

Altogether, these results are puzzling: on the one hand, they show that complexity accumulates despite selection (simple organisms being fitter than complex ones), and on the other hand, they suggest that complexity is not driven by a neutral process either. To disentangle the relationship between drift and selection, we analysed the evolution of complexity and fitness during the 250,000 generations of the experiments (Fig 5).

Figs. 5.A and 5.B illustrate that the dynamics are different for the two complexity measures but similar for the two environmental conditions. Indeed, in both complex and simple environments C_G quickly plateaus while C_P increases all along evolution (less obviously so for the highest mutation rate). Moreover, Fig. 5.C shows that, even though simple organisms are fitter than complex ones (see above), paradoxically, the latter improve their fitnesses while increasing in complexity. This shows that, in our experiments, complexity is driven by a strong sign-epistasis mechanism that, in the fitness landscape of Aevol, sets simple functional structures and complex ones apart. This sign-epistasis initiates a

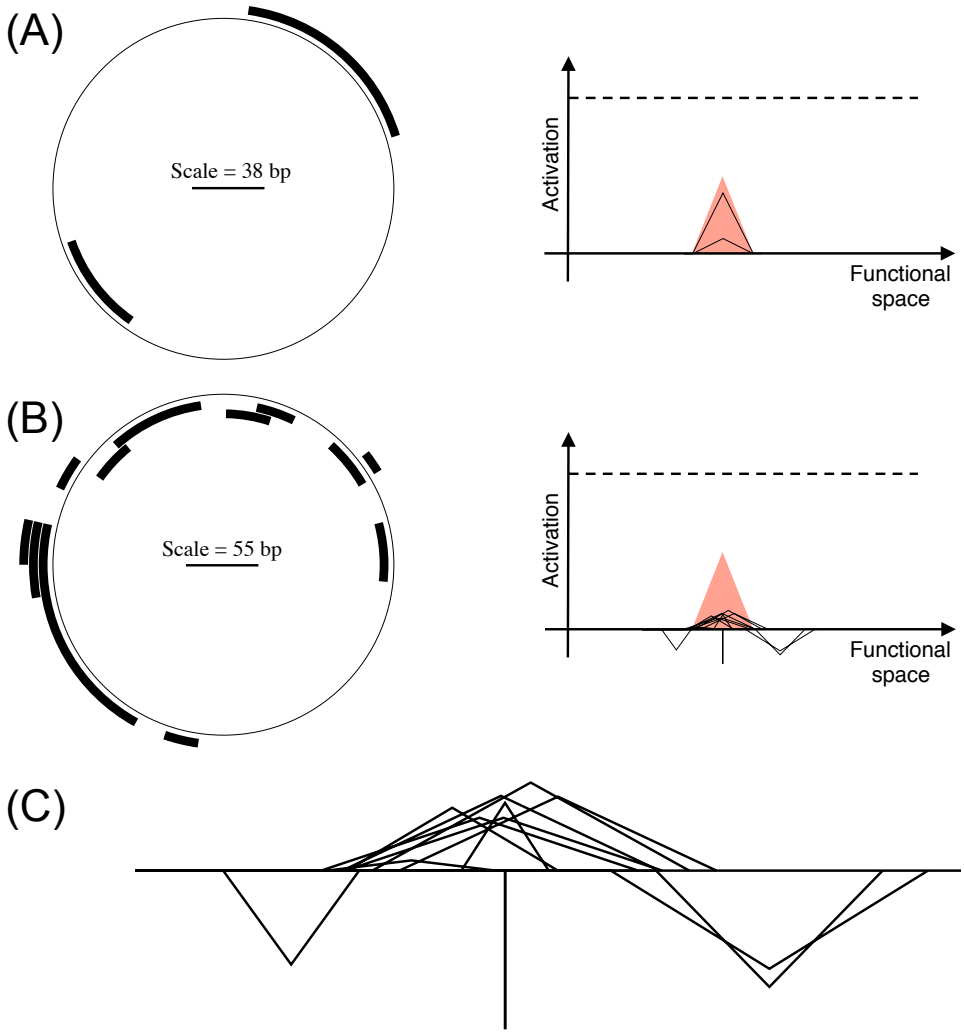

Figure 4: A simple (A) and a complex (B) organism. Both evolved for 250,000 generations under a mild mutation rate (10^{-5} mut.bp^{-1}.gen^{-1}) in simple conditions (triangular target). Left: circular genomes and genes (black arcs). Notice the non-coding sequences on both genomes and also the genes overlap on the genome of the complex organism (B). Right: proteins (black triangles) and phenotypic targets (red filled triangle). Panel (C) zooms on the protein structure of the complex organism. The simple organism has two genes, two proteins and a functional complexity $C_P = 4$. The complex organism has 12 genes, 11 functional proteins and $C_P = 24$.

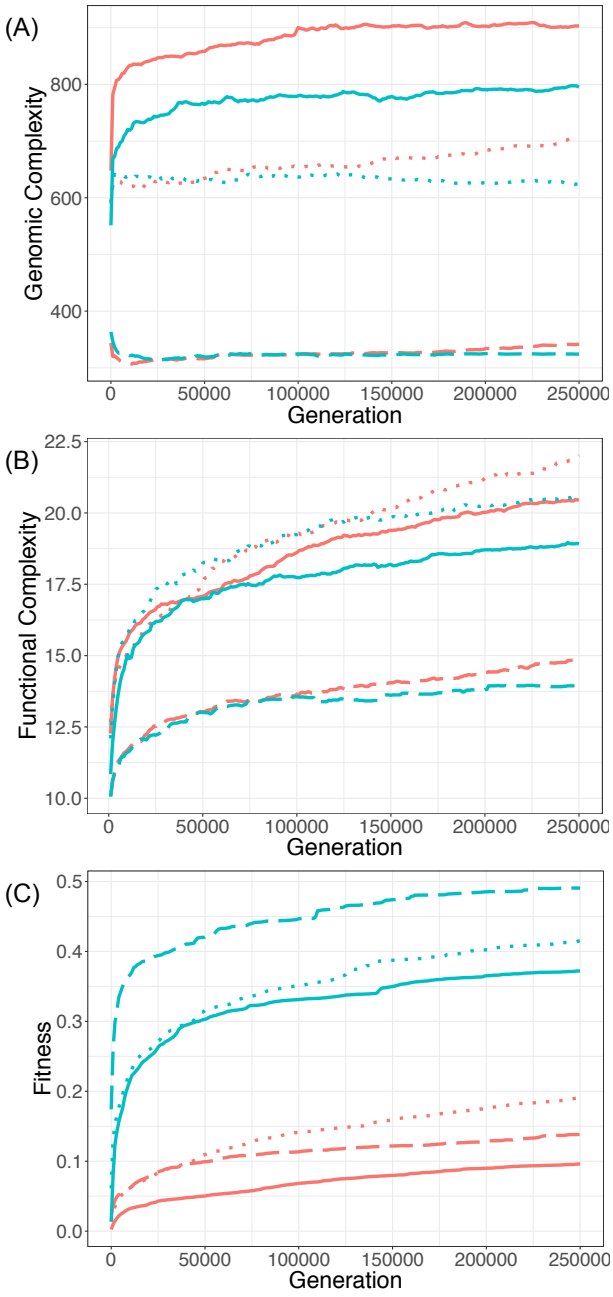

Figure 5: **(A)** Mean genomic complexity C_G along generations for organisms that evolved in the simple environment (blue lines) and in the complex environment (red lines) and for the three different mutation rates. None of the differences are statistically significant at generation 250,000 except the effect of a harsh mutation rate (p-value $< 10^{-4}$).
(B) Mean functional complexity C_P along generations for organisms that evolved in the simple environment (blue lines) and in the complex environment (red lines) and for the three different mutation rates. None of the differences are statistically significant at generation 250,000 except the effect of a harsh mutation rate (p-value $< 10^{-3}$).
(C) Mean fitness f along generations for organisms that evolved in the simple environment (blue lines) and in the complex environment (red lines) and for the three different mutation rates. The effect of mutation rates on fitness are all statistically significant at generation 250,000 in complex conditions and for the two extreme mutation rates in simple conditions. Plain line: $\mu = 10^{-6}$. Dotted line: $\mu = 10^{-5}$. Dashed line: $\mu = 10^{-4}$.

"complexity ratchet" (Liard et al, 2018, 2020) that pushes organisms toward greater complexity, meanwhile improving their fitness but staying far below the fitness of simple organisms.

This means that the negative complexity-fitness correlation observed in the simple environment is due to some initial contingent events ("frozen accidents"). Depending on their very first evolutionary steps, organisms initiate different evolutionary trajectories: some initiate a "gene optimisation" trajectory (typically through substitutions) while others initiate a "gene duplication-divergence" trajectory. In a complex environment, both strategies ultimately converge as complexity is required to adapt to the Gaussian target. But in a simple environment, the gene-optimization strategy and the gene-duplication strategy are antagonistic because of sign-epistasis. Hence, once organisms have set themselves on either trajectory, they can hardly switch to the other, and the longer the evolutionary history, the harder the switch becomes. We verified this theory by evolving populations without any chromosomal rearrangement mechanism. We observed that, in these new conditions, 98% of the simulations lead to simple organisms (to be compared to the 23.7% of simple organisms in presence of both point mutations and rearrangements), confirming that the complexity ratchet is indeed fuelled by duplications. This also shows that a minimum diversity of mutational operators is required for the ratchet to be effective and suggests that including such a variety of operators is mandatory to observe the whole complexity of the evolutionary process.

By evolving initially simple organisms in a simple environment, we were able to observe the effect of a "complexity ratchet". Now, by comparing the complexity levels in simple and complex environments, we can estimate the power of this ratchet. Indeed, we first showed that complexity is not higher in complex environments, showing that the ratchet is at least as powerful as direct selection for complexity. Second, we have observed an effect of the mutation rate on the complexity levels and on fitness levels (Fig. 5): high mutation rates strongly limit the maximum level of genomic complexity (Fig. 5.A).This suggests that evolution of complexity must be analysed in a multi-scale framework: the complexity ratchet drives complexity at the functional level but the functional level has to be encoded within the genomic sequence. Hence, mutational robustness, by bounding the amount of information a genome can store (Wilke et al, 2001; Knibbe et al, 2007), loosely bounds the functional complexity. Yet, at the functional level, complexity continuously increases (Fig. 5.B), driven by the complexity ratchet that slowly increases fitness (Fig. 5.C).

5 Conclusion

In this chapter, we have explored the relationship ESB maintains with complexity. We successively discussed the apparently low interest of ESB for the question of complexity. We then showed how complex models can be used to explore how evolving systems accumulate complexity. Finally, we presented an experiment that revealed the existence of a "complexity ratchet" fuelled by sign-epistasis (Liard et al, 2018, 2020). Within the field of ESB, this result deserves a specific discussion. Indeed, it shows that, when analysing a biological system, there may be no relationship between the complexity of the structure and that of the function. Because of the complexity ratchet, a simple function can very well be carried out by very complicated systems, even when simple solutions exist. Interestingly, sign-epistasis has recently been identified in signaling cascades (Nghe et al, 2018), a system that is well known for its unnecessary complexity (Soyer and Bonhoeffer, 2006).

On the methodological side, complex models raise a difficult question: given the complexity of these models, how can one ensure that the results (and, here, the complexity ratchet), are valid, i.e. transferable to "real" living systems but also to evolutionary biology and to systems biology that don't make use of this kind of models (and hence don't trust them). On that question, the answer has been given more than 20 years ago by Volker Grimm in an enlightening article (Grimm, 1999) "*The decisive thing with modelling is not the model per se, but what the model and working with the model does to our mind. [...] If the whole process of modelling has succeeded, something will have happened in our head, namely that an understanding of relationships has emerged. We should then be in a position to communicate our insights to others without referring to the model*". It has no meaning to discuss whether Aevol is "true" or "false"; But if we are able to explain the complexity ratchet in biological terms independently from the model that revealed it, then the question of the model's rightness no longer matters.

References

Adami C (2002) What is complexity? BioEssays 24(12):1085–1094

Adami C (2006) Digital genetics: unravelling the genetic basis of evolution. Nat Rev Genet 7(2):109

Adami C, Ofria C, Collier TC (2000) Evolution of biological complexity. PNAS 97(9):4463–4468

Albantakis L, Hintze A, Koch C, Adami C, Tononi G (2014) Evolution of integrated causal structures in animats exposed to environments of increasing complexity. PLoS Comput Biol 10(12):e1003966

Alberch P (1991) From genes to phenotype: dynamical systems and evolvability. Genetica 84(1):5–11

Batut B, Knibbe C, Marais G, Daubin V (2014) Reductive genome evolution at both ends of the bacterial population size spectrum. Nat Rev Microb 12(12):841

Beslon G, Parsons DP, Sanchez-Dehesa Y, Pena JM, Knibbe C (2010) Scaling laws in bacterial genomes: A side-effect of selection of mutational robustness? Biosystems 102(1):32–40

Clune J, Mouret JB, Lipson H (2013) The evolutionary origins of modularity. Proc Roy Soc B 280(1755):20122863

Crombach A, Hogeweg P (2008) Evolution of evolvability in gene regulatory networks. PLoS Comp Biol 4(7):e1000112

Darwin C (1859) On the Origin of Species by Means of Natural Selection. J. Murray, London

Dawkins R (1997) Human chauvinism. Evolution 51(3):1015–1020

De Visser JAG, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al (2003) Perspective: evolution and detection of genetic robustness. Evolution 57(9):1959–1972

Elliott TA, Gregory TR (2015) What's in a genome? the c-value enigma and the evolution of eukaryotic genome content. Phil Trans Roy Soc B 370(1678):20140331

Espinosa-Soto C, Wagner A (2010) Specialization can drive the evolution of modularity. PLoS Comp Biol 6(3)

Gould SJ (1996) Full House: The Spread of Joy from Plato to Darwin. Harmony Books

Greene CS, Tan J, Ung M, Moore JH, Cheng C (2014) Big data bioinformatics. J Cell Phys 229(12):1896–1900

Grimm V (1999) Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecological modelling 115(2-3):129–148

Hindré T, Knibbe C, Beslon G, Schneider D (2012) New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nat Rev Microb 10(5):352

Kashtan N, Alon U (2005) Spontaneous evolution of modularity and network motifs. PNAS 102(39):13773–13778

Knibbe C, Coulon A, Mazet O, Fayard JM, Beslon G (2007) A long-term evolutionary pressure on the amount of noncoding DNA. Mol Biol Evol 24(10):2344–2353

Lenski RE, Ofria C, Pennock RT, Adami C (2003) The evolutionary origin of complex features. Nature 423(6936):139–144

Liard V, Parsons D, Rouzaud-Cornabas J, Beslon G (2018) The complexity ratchet: Stronger than selection, weaker than robustness. In: Ikegami T, Virgo N, Witkowski O, Oka M, Suzuki R, Iizuka H (eds) Artificial Life Conference Proceedings, MIT Press, pp 250–257

Liard V, Parsons DP, Rouzaud-Cornabas J, Beslon G (2020) The complexity ratchet: Stronger than selection, stronger than evolvability, weaker than robustness. Artificial Life 26:38–57

Loewe L (2016) Systems in evolutionary systems biology. Encyclopedia of Evolutionary Biology 4:297–318

Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW (2011) How a neutral evolutionary ratchet can build cellular complexity. IUBMB life 63(7):528–537

Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. PNAS 104(suppl 1):8597–8604

Maslov S, Krishna S, Pang TY, Sneppen K (2009) Toolbox model of evolution of prokaryotic metabolic networks and their regulation. PNAS 106(24):9743–9748

Maynard Smith J, Szathmary E (1997) The major transitions in evolution. Oxford University Press

Mayr E (1961) Cause and effect in biology. Science 134(3489):1501–1506

McShea DW (1996) Metazoan complexity and evolution: Is there a trend? Evolution 50(2):477–492

McShea DW, Brandon RN (2010) Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems. University of Chicago Press

Moran NA (2007) Symbiosis as an adaptive process and source of phenotypic complexity. PNAS 104(suppl 1):8627–8633

Nghe P, Kogenaru M, Tans SJ (2018) Sign epistasis caused by hierarchy within signalling cascades. Nat Comm 9(1):1–9

O'Neill B (2003) Digital evolution. PLoS Biol 1(1):e18+

Pavlicev M, Wagner GP (2012) A model of developmental evolution: selection, pleiotropy and compensation. TREE 27(6):316–322

Pigliucci M (2008) Is evolvability evolvable? Nat Rev Genet 9(1):75

Pigliucci M (2009) The end of theory in science? EMBO reports 10(6):534–534

Ray TS (1991) An approach to the synthesis of life. Artificial life II pp 371–408

Rutten JP, Hogeweg P, Beslon G (2019) Adapting the engine to the fuel: mutator populations can reduce the mutational load by reorganizing their genome structure. BMC Evol Biol 19(1):191

Simon HA (1962) The architecture of complexity. Proc Am Phil Soc 106(6):467–482

Soyer OS, Bonhoeffer S (2006) Evolution of complexity in signaling pathways. PNAS 103(44):16337–16342

Soyer OS, O'Malley MA (2013) Evolutionary systems biology: what it is and why it matters. BioEssays 35(8):696–705

Thomas Jr CA (1971) The genetic organization of chromosomes. Ann Rev Genet 5(1):237–256

Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412(6844):331

Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE (2011) Second-order selection for evolvability in a large escherichia coli population. Science 331(6023):1433–1436

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proc. sixth international congress of Genetics, vol 1, pp 356–366

Zaman L, Meyer JR, Devangam S, Bryson DM, Lenski RE, Ofria C (2014) Coevolution drives the emergence of complex traits and promotes evolvability. PLoS Biol 12(12):e1002023